(2)

Asymptotic Distribution of the Likelihood Ratio Test That a Mixture of Two Binomials is a Single Binomial

Herman Chernoff
Department of Statistics
Harvard University

and

Eric Lander
Whitehead Institute
Massachusetts Institute of Technology

Technical Report No. ONR-C-7

May, 1991

Asymptotic Distribution of the Likelihood Ratio Test That a Mixture of Two Binomials is a Single Binomial

Herman Chernoff
Department of Statistics
Harvard University

and

Eric Lander
Whitehead Institute
Massachusetts Institute of Technology

ABSTRACT

A problem of interest in genetics is that of testing whether a mixture of two binomial distributions $B_i(k,p)$ and $B_i(k,1/2)$ is simply the pure distribution $B_i(k,1/2)$. This problem arises in determining whether we have a genetic marker for a gene responsible for a *heterogeneous trait*, that is a trait which is caused by any one of several genes. In that event we would have a nontrivial mixture involving $0 < p < 0.5$ where p is a recombination probability.

Standard asymptotic theory breaks down for such problems which belong to a class of problems where a *natural* parameterization represents a single distribution, under the hypothesis to be tested, by infinitely many possible parameter points. That difficulty may be eliminated by a transformation of parameters. But in that case a second problem appears. The regularity conditions demanded by the applicability of the Fisher Information fails when $k > 2$. We present an approach where use is made of the Kullback Leibler information, of which the Fisher information is a limiting case.

Several versions of the binomial mixture problem are studied. The asymptotic analysis is supplemented by the results of simulations. It is shown that as $n \to \infty$, the asymptotic distribution of twice the logarithm of the likelihood ratio corresponds to the square of the supremum of a Gaussian stochastic process with mean 0, variance 1 and a well behaved covariance function. As $k \to \infty$ this limiting distribution grows stochastically as the square root of $\log k$.

AMS 1980 *Subject Classifications* Primary 62F05, 62P10; Secondary 60G15

Keywords and Phrases asymptotic distribution, likelihood ratio, extreme values, mixture, genetic marker, heterogeneous trait.

1. Introduction.

A problem of interest in genetics is that of testing the hypothesis that a mixture of two binomial distributions $Bi(k,p)$ and $Bi(k,1/2)$ is the degenerate case of the single binomial $Bi(k,1/2)$. This problem arises in determining whether there is a marker for a genetically heterogeneous trait, i.e., a trait that can be caused by a mutation at any one of several different loci. For parametric hypothesis testing problems it is customary to use the generalized likelihood ratio as a test statistic. Under standard regularity conditions, a classical result of Wilks (1938) states that if the hypothesis is true, twice the logarithm of the likelihood ratio has, asymptotically, a chi-square distribution.

The regularity conditions are not satisfied for our mixture problem. Moreover, under the parametrization ordinarily used for this type of problem, the hypothesis, which is simple and uniquely determines the distribution of the data, corresponds to an infinite set of parameter points designating the mixture fraction and the probability p. This complication may be eliminated by introducing an alternative parametrization.

For the case $k = 2$ with this reparametrization, most of the regularity conditions are satisfied, and a generalization of the Wilks result (Chernoff, 1954) establishes that the asymptotic distribution of twice the logarithm of the likelihood ratio is a mixture of three distributions, two of which are those of chi-square with one and two degrees of freedom. However, for $k > 2$, the regularity conditions no longer apply.

For this special problem, the distribution of the likelihood ratio can be determined by simulation. However, asymptotic theory is useful in understanding generalizations of our problem. One generalization is for the model where independent observations correspond to different values of k. Another is when we wish to test that a mixture of $Bi(k,p_1)$ and $Bi(k,p_2)$ with unknown p_1 and p_2 is really a single binomial. We shall analyze the first of these generalizations. These problems belong to a large class of problems where regularity conditions fail, and include mixture problems discussed by Ghosh and Sen (1985) and change point and segmented regression problems treated by Feder (1975).

The main idea behind our approach is that the Fisher Information which characterizes the behavior of the maximum likelihood estimate (MLE) degenerates in problems which lack regularity. However, the Fisher Information is a limiting case of the Kullback-Leibler (KL) Information, which is the expectation of the likelihood ratio, and is the natural measure of the ability to use data to discriminate between alternative hypotheses. The study of the KL Information for nearby alternatives clarifies what constitutes appropriate parametrizations and the asymptotic behavior of the likelihood ratio as well as the MLE.

In our problems we shall express the asymptotic distribution of the logarithm of the likelihood ratio in terms of the maximum of the square of a relatively simple Gaussian stochastic process.

In Section 2 we present formal statements of several problems and the appropriate parametrization. In Section 3, we discuss the asymptotic distribution under the null hypothesis for the case of $k = 2$. In Section 4, the asymptotic distribution for the case of arbitrary k is derived. In Section 5, extensions of these results are presented to include several values of k, a restricted version of the problem where $p > 1/2$, noncentral results and large deviation results. Derivations appear in an appendix. Section 6 presents results of simulations comparing asymptotic and finite sample distributions.

The main asymptotic result is that under the null hypothesis, twice the logarithm of the likelihood ratio behaves like the square of the maximum of a stochastic process in a variable ϕ, and for each value of ϕ, the process has mean 0 and variance 1. The problem discussed here is a special case of a more general theory which applies to mixture problems and change point problems. The most general discussion to data, one which applies to mixture problems and would include our problem as a special case, is due to Ghosh and Sen (1985).

2. Problem Statements and Reparametrization.

We present a formal statement of the problem and several variations. In what follows $\mathcal{L}(X)$ stands for the distribution (law) of X and $\mathcal{L}(X|Y)$ is the conditional distribution of X given Y. The distribution and expectation for a given value of the parameter θ are represented by $\mathcal{L}_\theta$ and E_θ. The binomial distribution corresponding to k trials with probability p is designated by $Bi(k,p)$ and $N(\mu,\Sigma)$ represents the normal distribution with mean μ and covariance matrix Σ. The chi-square distribution with m degrees of freedom is written $\mathcal{L}(\chi^2_m)$.

PROBLEM 1. Let $X_1, X_2, ..., X_n$ be *i.i.d* observations on a random variable X for which the distribution

$$\mathcal{L}(X) = \alpha Bi(k,p) + (1-\alpha)Bi(k,1/2) \tag{2.1}$$

where α and p are unknown and $0 \le \alpha \le 1$ and $0 \le p \le 1$. The hypothesis, $H_o : p = 1/2$ or $\alpha = 0$, is tested using the likelihood ratio test. Assuming H_o is true what are the distributions of the likelihood ratio and the maximum likelihood estimates of α and p?

PROBLEM 2. Let $X_1, X_2, ..., X_n$ be independent observations where

$$\mathcal{L}(X_i) = \alpha Bi(k_i,p) + (1-\alpha)Bi(k_i,1/2) ,$$

the k_i are known, $n_k = n\lambda_k$ is the number of times $k_i = k$, $\Sigma\lambda_k = 1$, $0 \le p \le 1$ and $0 \le \alpha \le 1$. The hypothesis, $H_o : p = 1/2$ or $\alpha = 0$, is tested as in Problem 1.

PROBLEM 3. Consider the variations of Problems 1 and 2 where p is restricted by $0 \le p \le 0.5$. These variations are those most relevant for the application to genetics. There, p represents the recombination fraction of a proposed marker to one of the loci for the trait and achieves a maximum value of $1/2$ when there is no linkage. Furthermore k represents the size of the family studied and may vary from one family to another. We will refer to these variations as the *one sided* cases.

PROBLEM 4. Consider the variation of Problem 1 where

$$\mathcal{L}(X) = \alpha Bi(k, p_1) + (1 - \alpha)Bi(k, p_2)$$

with p_1, p_2 and α unknown, $0 \leq p_1 \leq 1$, $0 \leq p_2 \leq 1$, $0 \leq \alpha \leq 1$. It is desired to test the hypothesis $H_o : p_1 = p_2$ or $\alpha = 0$ or 1.

Under H_o in Problem 1, $\mathcal{L}(X) = Bi(k, 1/2)$. The same distribution applies when $p = 1/2$, no matter what the value of α is, or when $\alpha = 0$, no matter what the value of p. In effect the hypothesis H_o, which corresponds to $\{(\alpha, p) : \alpha = 0, 0 \leq p \leq 1\} \cup \{(\alpha, p) : p = 1/2, 0 \leq \alpha \leq 1\}$, really corresponds to only one point in the space of distributions. Thus a more "natural" parametrization should have only one point correspond to the above set. We offer, as an alternative, the following parametrization which will prove convenient. Let $\theta = (\theta_1, \theta_2)^T$ (the exponent T is used for transpose) where

$$\begin{aligned} \theta_1 &= \frac{1}{2}\alpha(p - 1/2) \\ \theta_2 &= \alpha(p - 1/2)^2 \end{aligned} \tag{2.2}$$

then

$$\alpha = 4\theta_1^2/\theta_2 \tag{2.3}$$

and

$$p = \frac{1}{2}(1 + \frac{\theta_2}{\theta_1}) = (1 + \phi)/2 \tag{2.4}$$

where

$$\phi = \theta_2/\theta_1 = 2p - 1 \ . \tag{2.5}$$

In the new parametrization, H_o corresponds to $\theta = 0$ or, equivalently, $\theta_1 = \theta_2 = 0$. The range of (θ_1, θ_2) lies between the line $\theta_2 = \theta_1$ and the parabola $\theta_2 = 4\theta_1^2$ for $0 \leq \theta_1 \leq 1/4$ and between $\theta_2 = -\theta_1$ and $\theta_2 = 4\theta_1^2$ for $-1/4 \leq \theta_1 \leq 0$. Note that $\theta_2 \geq 0$ and ϕ ranges from -1 to 1, and has the same sign as θ_1.

In terms of the new parametrization, the probability density of X is

$$f(x,\theta) = 2^{-k}\binom{k}{x}\left\{\frac{4\theta_1^2}{\theta_2}\left(1+\frac{\theta_2}{\theta_1}\right)^x\left(1-\frac{\theta_2}{\theta_1}\right)^{k-x} + \left(1-\frac{4\theta_1^2}{\theta_2}\right)\right\}, \quad x = 0,1,\ldots,k$$

and

$$f(x,0) = 2^{-k}\binom{k}{x}, \quad x = 0,1,\ldots,k .$$

The likelihood ratio is

$$v(X,\theta) = 1 + u(X,\theta) = \frac{f(X,\theta)}{f(X,0)} = 1 + \frac{4\theta_1^2}{\theta_2}\left\{\left(1+\frac{\theta_2}{\theta_1}\right)^X\left(1-\frac{\theta_2}{\theta_1}\right)^{k-X} - 1\right\} \tag{2.6}$$

We may write

$$u(x,\theta) = 4\theta_1 P_{kx}(\phi) \tag{2.7}$$

where

$$P_{kx}(\phi) = \frac{(1+\phi)^x(1-\phi)^{k-x} - 1}{\phi} \tag{2.8}$$

is a polynomial of degree $k-1$ in ϕ.

The logarithm of the likelihood is

$$\ell n(f(X,\theta)/f(X,0)] = \ell n[1 + u(X,\theta)].$$

The Fisher Information is defined by

$$J(\theta) = E_\theta\left\{\left[\frac{\partial \ell n f(X,\theta)}{\partial\theta}\right]\left[\frac{\partial \ell n f(X,\theta)}{\partial\theta}\right]^T\right\} \tag{2.9}$$

where $\partial/\partial\theta$ represents the column vector whose components are the partial derivatives with respect to the components of θ. The Kullback Leibler Information is

$$K(\theta,\theta^*) = E_\theta\{\ell n[f(X,\theta)/f(X,\theta^*)]\}. \tag{2.10}$$

3. Asymptotic distribution of likelihood ratio when $k = 2$.

For Problem 1, the case $k = 2$ becomes relatively simple. There, we have

$$\begin{aligned} f(0,\theta) &= 0.25 - 2\theta_1 + \theta_2 \\ f(1,\theta) &= 0.5 - 2\theta_2 \\ f(2,\theta) &= 0.25 + 2\theta_1 + \theta_2 \end{aligned} \tag{3.1}$$

We observe data on a multinomial distribution with three probabilities depending linearly on two parameters. The Fisher Information is easily calculated and

$$J(0) = \begin{pmatrix} 32 & 0 \\ 0 & 16 \end{pmatrix} \tag{3.2}$$

Let $\hat{\theta}_u$ be the MLE unrestricted by the restrictions on the range of θ. Standard theory tells us that when $\theta = 0$, $\mathcal{L}_0(\sqrt{n}\hat{\theta}_u) \to N(0, J^{-1}(0))$. Applying Chernoff (1954) with the restriction on θ, it follows that $L = \ell n$ (likelihood) satisfies

$$\mathcal{L}_0(2L) \to \frac{1}{2}\mathcal{L}(\chi_1^2) + 2\lambda \cdot \mathcal{L}(\chi_2^2) + \left(\frac{1}{2} - 2\lambda\right)\mathcal{L}(Y_1^2 | 0 < Y_2 < \sqrt{2}Y_1) \quad \text{as} \quad n \to \infty, \tag{3.3}$$

where $\mathcal{L}(\chi_m^2)$ is the chi-square distribution with m degrees of freedom, Y_1 and Y_2 are independent $N(0,1)$ random variables, and

$$\lambda = \frac{1}{2\pi} \arctan(1/\sqrt{2}) = 0.098 \tag{3.4}$$

Alternatively, we may write

$$P_0\{2L < x\} \to \left[\Phi(\sqrt{x}) - \frac{1}{2}\right] + 2\lambda\left[1 - e^{-x/2}\right] + 2\int_0^{\sqrt{x}} \phi(t)\left[\Phi(\sqrt{2}t) - \frac{1}{2}\right] dt \tag{3.5}$$

where ϕ and Φ are the density and *cdf* for the $N(0,1)$ distribution. Some detail is presented in Appendix 1.

In Problem 3, a similar analysis involves a further restriction on θ. There, θ is restricted to the right half of the range of θ for Problem 1. Here

$$\mathcal{L}_0(2L) \to \frac{1}{2}\mathcal{L}(\chi_1^2) + \lambda\mathcal{L}(\chi_2^2) + \left(\frac{1}{2} - \lambda\right)\mathcal{L}(0) \tag{3.6}$$

where $\mathcal{L}(0)$ is the distribution which attaches probability one to the value 0.

4. Asymptotic distribution of likelihood ratio for arbitrary k.

By relating the likelihood ratio to its expectation, the Kullback Leibler Information, we shall show that for specified $\phi = \theta_2/\theta_1$, twice the logarithm of the likelihood ratio is simply expressed in terms of $S(Z,\phi)$ where $S(Z,\phi)$ is a polynomial in ϕ, linear in Z, and Z is an asymptotically normal vector random variable. The resulting characterization of the distribution of the likelihood ratio involves maximizing with respect to ϕ.

4.1 The Kullback Leibler Information.

First let us evaluate the KL Information based on a single observation for Problem 1. Let

$$U = u(X,\theta) = \frac{4\theta_1}{\phi}\left\{(1+\phi)^X(1-\phi)^{k-X} - 1\right\} = 4\theta_1 P_{Kx}(\phi) \tag{4.1}$$

Then, for $\theta = o(1)$,

$$K(0,\theta) = E_o\{-\ell n(1+U)\} = E_o\left\{-U + \frac{1}{2}U^2\right\} + o(\theta_1^2)$$

But we can see, without calculating that

$$E_o(U) = E_o\left[\frac{f(X,\theta)}{f(X,0)} - 1\right] = 0, \tag{4.2}$$

and hence

$$E_o\left\{(1+\phi)^X(1-\phi)^{k-X}\right\} = 1.$$

However, to evaluate $E_o(U^2)$, we must calculate

$$E_o\left\{(1+\phi)^{2X}(1-\phi)^{2(k-X)}\right\} = \sum_{x=0}^{k} 2^{-k}\binom{k}{x}\left[(1+\phi)^{2x}(1-\phi)^{2(k-x)}\right]$$
$$= \left[\frac{(1+\phi)^2 + (1-\phi)^2}{2}\right]^k = (1+\phi^2)^k,$$

Thus

$$E_o(U^2) = \frac{16\theta_1^2}{\phi^2}\Big[(1+\phi^2)^k - 1\Big] = 16\theta_1^2 P_{kk}(\phi^2)$$

And

$$K(0,\theta) = 8\theta_1^2 P_{kk}(\phi^2) + o(\theta_1^2). \tag{4.3}$$

Applying the additivity of KL Information, we have

LEMMA 1. *For Problem 1, the KL Information* $K(0,\theta)$ *is* $8n[\theta_1^2 P_{kk}(\phi^2) + o(\theta_1^2)]$. *For Problem 2, it is* $8n[\theta_1^2 \Sigma\lambda_K p_{kk}(\phi^2) + o(\theta_1^2)]$.

The form of $K(0,\theta)$ suggests the importance of θ_1 in our reparametrization. For $k = 2, K(0,\theta) \approx 16\theta_1^2 + 8\theta_2^2 = \frac{1}{2}\theta^T J(0)\theta$. However, for $k = 3$, $K(0,\theta) \approx 24\theta_1^2 + 24\theta_2^2 + 8\theta_2^4/\theta_1^2$ which does not behave well as θ_1 and θ_2 approach zero. The lack of regularity for $k = 3$ is associated with this "poor behavior."

4.2 The likelihood ratio for Problem 1.

Let us assume in Problem 1 that M_j of the n observations have $X_i = j$. Then $M = (M_0, M_1, \ldots, M_k)^T$ has a multinomial distribution, and from the central limit theorem it follows that $M_j = nf(j,0) + \sqrt{n}Z_j$ where

$$\mathcal{L}_o(Z) \to N(0,\Sigma) \quad \text{as} \quad n \to \infty \tag{4.4}$$

and

$$\Sigma = \|\sigma_{ij}\| = \|f(i,0)\delta_{ij} - f(i,0)f(j,0)\|, \quad i,j = 0,1,\ldots,k \tag{4.5}$$

with $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The logarithm of the likelihood ratio is $L = \sup L(\theta)$ where, for $\theta = 0(n^{-1/2})$

$$\begin{aligned} L(\theta) &= \Sigma\Big\{nf(j,0) + \sqrt{n}Z_j\Big\}\ell n[1 + u(j,\theta)] \\ &= \Sigma\Big\{nf(j,0)[u(j,\theta) - \frac{1}{2}u^2(j,\theta)] + \sqrt{n}Z_j u(j,\theta)\Big\} + o_p(1)\cdot \\ &= -8\theta_1^2 P_{kk}(\phi^2)\cdot n + 4\sqrt{n}\theta_1 S(Z,\phi) + o_p(1) \end{aligned}$$

where

$$S(Z,\phi) = \Sigma Z_J p_{kj}(\phi). \tag{4.6}$$

For fixed ϕ, the maximum of the main terms of $L(\theta)$ is attained at

$$\theta_1 = S(Z,\phi)/4\sqrt{n}P_{kk}(\phi^2). \tag{4.7}$$

However, since $\theta_2 > 0$, θ_1 is restricted to have the same sign as ϕ, and the maximum over this restricted range is $\frac{1}{2}T^2(\phi)$ where

$$T(\phi) = \max\left[0, \frac{\text{sgn}(\phi)S(Z,\phi)}{\sqrt{P_{kk}(\phi^2)}}\right] \tag{4.8}$$

Thus

$$2L = \sup_{|\phi|\leq 1} T^2(\phi) + o_p(1). \tag{4.9}$$

and we have established

THEOREM 1. *For problem 1, the asymptotic distribution of twice the logarithm of the likelihood ratio is that of* $\sup_{|\phi|\leq 1} T^2$ *where* $T(\phi)$ *is defined in* (4.8) *with* $\mathcal{L}_o(Z)$ *replaced by* $N(0,\Sigma)$.

4.3 The Stochastic process $S(Z,\phi)$.

In Problem 1, the limiting distribution of L under H_o is determined by the fact that $\mathcal{L}_o(Z) \to N(0,\Sigma)$ as $n \to \infty$. By a continuity argument, this limiting distribution may be obtained by assuming, as we shall in this subsection, that $\mathcal{L}(Z) = N(0,\Sigma)$. Then the stochastic process $S(Z,\phi)$ is a Gaussian process which may be expressed as a polynomial of degree $k-1$ in ϕ, i.e.,

$$S(Z,\phi) = \sum_{j=0}^{k} Z_J p_{kj}(\phi) = \sum_{i=1}^{k} W_i \phi^{i-1}. \tag{4.10}$$

The distribution of S is described in

PROPOSITION 1. *The coefficients W_i are independent normal random variables with mean 0 and variance $\binom{k}{i}$. The process $S(Z,\phi)$ has mean 0 variance $P_{kk}(\phi^2)$ and autocorrelation*

$$\rho(\phi_1,\phi_2) = P_{kk}(\phi_1\phi_2)/\Big[P_{kk}(\phi_1^2)P_{kk}(\phi_2^2)\Big]^{1/2}. \tag{4.11}$$

Proof: The W_i are linear functions of the Z_j and hence have a multivariate normal distribution with mean 0 . Also $E(S(Z,\phi))=0$. Then

$$E[S(Z,\phi_1)S(Z,\phi_2)] = \sum_{i,j=0}^{k} P_{ki}(\phi_1)\sigma_{ij}P_{kj}(\phi_2) = \sum_{i,j=1}^{k} \phi_1^{i-1}E(W_I w_j)\phi_2^{j-1} \tag{4.12}$$

But

$$\begin{aligned}
\sum_{i,j} P_{ki}(\phi_1)\sigma_{ij}P_{kj}(\phi_2) &= \sum_i P_{ki}(\phi_1)P_{ki}(\phi_2)f(i,0) \\
&\quad - \Big[\sum_i P_{ki}(\phi_1)f(i,0)\Big]\Big[\sum_j P_{kj}(\phi_2)f(j,0)\Big] \\
\sum_i P_{ki}(\phi)f(i,0) &= \sum_i 2^{-k}\binom{k}{i}\Big[\frac{(1+\phi)^i(1-\phi)^{k-i}-1}{\phi}\Big] = 0 \\
\sum_i P_{ki}(\phi_1)P_{ki}(\phi_2)f(i,0) &= \sum 2^k\binom{k}{i}\Big[\frac{(1+\phi_1)^i(1-\phi_1)^{k-i}-1}{\phi_1}\Big] \\
&\quad \times \Big[\frac{(1+\phi_2)^i(1-\phi_2)^{k-i}-1}{\phi_2}\Big] \\
&= \frac{1}{\phi_1\phi_2}\Big\{\Big[\frac{(1+\phi_1)(1+\phi_2)+(1-\phi_1)(1-\phi_2)}{2}\Big]^k - 1\Big\} \\
&= P_{kk}(\phi_1\phi_2)
\end{aligned}$$

The representation (4.11) for the autocorrelation follows. Equating coefficients in (4.12), it follows that $EW_I w_j = 0$ if $i \neq j$ and $EW_i^2 = \binom{k}{i}$.

One consequence of Proposition 1, is that the stochastic process $S(Z,\phi)/\sqrt{P_{kk}(\phi^2)}$, which is so important in $T(\phi)$, is (asymptotically) a Gaussian process with mean 0 and variance 1 for each value of ϕ.

5. Generalization, extensions and comments.

In this section we point out that the results for Problem 1 extend easily to Problems 2 and 3. These results are associated with the behavior of the maximum likelihood estimates of the parameters of the model which resembles that of easily derived the moment method estimates. A geometric interpretation of these results is presented.

Problem 4 may be treated along similar lines requiring a more complex analysis which will not be presented here.

Results are presented for the noncentral case where the hypothesis is almost true and for the restricted problem where the alternative hypothesis restricts p to be substantially removed from 0.5. The latter problem is relevant for the genetic application where we expect a strong linkage, if any, between the marker and one of the loci involved.

Finally, it is pointed out that the logarithm of the likelihood approaches infinity as $k \to \infty$. The rate of convergence is very slow, of order of magnitude of $(\log k)^{1/2}$.

5.1 Problems 2 and 3.

A straightforward extension of the argument for Problem 1, yields

THEOREM 2. *If* $n_k/n \to \lambda_k$ *as* $n \to \infty$ it with $\Sigma\lambda_k = 1$, *the limiting distribution of* $2L$ *under* H_o *in Problem 2 is* $\mathcal{L}\left[\sup_{|\phi|\le 1}\{T^2(\phi)\}\right]$ *where*

$$T(\phi) = \max\left[0, \frac{\operatorname{sgn}(\phi)\sum_k \lambda_k^{1/2}\sum_{i=1}^k W_{ki}\phi^{i-1}}{\sqrt{P(\phi^2)}}\right] \tag{5.1}$$

and $P(\phi^2) = \sum \lambda_K p_{kk}(\phi^2)$ *and the* W_{ki} *are independent* $N\left(0, \binom{k}{i}\right)$ *random variables.*

For Problem 3, the only change in the results for Problems 1 and 2 is that the supremum of $T^2(\phi)$ should be taken over the domain $-1 \le \phi \le 0$ which corresponds to $0 \le p \le 0.5$.

Problem 4 is more complicated to handle because it involves 3 parameters. We will not elaborate on it here, but it is subject to a similar analysis where the first stage of the maximization of the likelihood is with $p_1 - p_2$ kept fixed in place of ϕ.

5.2 Maximum likelihood and moment method estimates.

Returning to Problem 1, the results are associated with an implicit characterization of the MLE estimates. Here $\hat{\phi}$ is the value of ϕ which maximizes $T^2(\phi)$,

$$\hat{\theta}_1 = \frac{1}{4\sqrt{n}} \frac{S(Z,\hat{\phi})}{P_{kk}(\hat{\phi}^2)} + o_p(n^{-1/2})$$

and

$$\mathcal{L}_o(\sqrt{n}\hat{\theta}_1) \approx \mathcal{L}\Big[\frac{\operatorname{sgn}(\hat{\phi})T(\hat{\phi})}{4\sqrt{P_{kk}(\hat{\phi}^2)}}\Big]$$

While $\hat{\theta}_1 = O_p(n^{-1/2})$, the estimate $\hat{\phi}$ varies between -1 and +1. This is not especially surprising since ϕ is not *identified* when H_o holds.

Appendix 2 presents a derivation of the asymptotic properties of the unrestricted moment method estimator $\tilde{\theta}_u$ of θ. This derivation is relatively simple and the results were useful originally in clarifying the situation in Problem 1. It is seen that $\mathcal{L}_o(\sqrt{n}\tilde{\theta}_u) = N(0,\Sigma_u)$ where Σ_u is a diagonal matrix with diagonal entries $(16k)^{-1}$ and $(8k(k-1))^{-1}$. It follows that the $\tilde{\phi}_u = \tilde{\theta}_{2u}/\tilde{\theta}_{1u}$ has a limiting Cauchy distribution.

There does not seem to be a standard convention for modifying $\tilde{\theta}_u$ to $\tilde{\theta}$ in order to conform with the restriction. It seems natural to use some sort of projection, in which case $\tilde{\phi}$ would behave like the mixture of a truncated Cauchy with a probability of 0.5 at $\phi = 0$.

5.3 A geometric interpretation.

Appendix 1 and Section 4 present the results for $k = 2$ in Problem 1 in different ways. A geometric interpretation of the results of Section 4 relate these. The expression

$$T_1(\phi) = \operatorname{sgn}(\phi)\frac{S(Z,\phi)}{\sqrt{P_{kk}(\phi^2)}} = \operatorname{sgn}(\phi)\frac{\sum_{i=1}^{k} W_i\phi^{i-1}}{\sqrt{P_{kk}(\phi^2)}}$$

can also be written

$$T_1(\phi) = \sum_{i=1}^{k} Y_i g_i(\phi) = Y^T g/\sqrt{g^T g}$$

where the components of g are

$$g_i(\phi) = \text{sgn}(\phi)\binom{k}{i}^{1/2}\phi^{i-1}, \qquad i = 1, 2, \ldots, k$$

and $\mathcal{L}(Y) = N(0, I)$. Thus if $T_1(\phi)$ is positive, it may be interpreted as the length of the projection of Y onto the ray from the origin through g.

Consider the "cone" made up of all of these rays as ϕ varies from -1 to +1. This cone is a two dimensional surface in k dimensional space. For $k = 2$, that surface is simply the angle between the horizontal axis and the line with slope $1/\sqrt{2}$ and the reflection of that angle about the vertical axis. Our asymptotic expression for T^2 is the squared length of the projection of Y onto this surface. The MLE $\hat{\phi}$ corresponds to the ray on which the projection falls and $\sqrt{n}\hat{\theta}_1$ is the ratio of the length of the projection to four times the length of $g(\hat{\phi})$.

5.4 The noncentral case.

In regular problems, the limiting chi-square distribution of the Wilks result becomes a noncentral chi-square when the hypothesis is false but the true value of the parameter is close to the set of parameter values under which the hypothesis is true. In Appendix 3 we derive the following analogous result for Problem 1.

THEOREM 3. *In Problem 1 let the true value* θ_{0n} *of the parameter be* $n^{-1/2}(\theta_1^*, \theta_2^*)^T$ *for fixed* θ_1^*, *and* $\theta_2^* = \phi^*\theta_1^*$. *Then*

$$\mathcal{L}_{\theta_{0n}}(2L) \to \mathcal{L}\left\{\sup_{|\phi|\le 1} T^{*2}(\phi)\right\}$$

where

$$T^*(\phi) = \max\left[0, \frac{\text{sgn}(\phi)S^*(Z,\phi)}{\sqrt{P_{kk}(\phi^2)}}\right],$$

$$S^*(Z,\phi) = \sum_{i-1}^{k} W_i^*\phi^{i-1},$$

and the W_i^* are independent with

$$\mathcal{L}(W_i^*) = N\left(4\binom{k}{i}\theta_1^*\phi^{*(i-1)}, \binom{k}{i}\right).$$

When the true value of θ is θ_0 which is substantially different from 0, we have what may be called the *large deviation case.* Then the distribution of the logarithm of the likelihood behaves like $N(Nk(\theta_0,0), nV(\theta_0))$ where

$$K(\theta_0,0) = E_{\theta_0}\left\{\ell n[f(X,\theta_0)/f(X,0)]\right\}$$

and

$$V(\theta_0) = \mathrm{Var}_{\theta_0}\left\{\ell n[f(X,\theta_0)/f(X,0)]\right\}.$$

5.5 The restricted problem.

If we restrict p to be in the interval $0 \le p \le p^* < 0.5$, then $-1 \le \phi \le \phi^* = 2p^* - 1 < 0$, and we have

$$\mathcal{L}_0(2L) \to \mathcal{L}\left\{\max\left[0, \sup_{-1\le\phi\le\phi^*}\left(\frac{-S(Z,\phi)}{\sqrt{P_{kk}(\phi^2)}}\right)\right]^2\right\}$$

For the special case $k = 2$, the analysis of Appendix 1 is easily extended to yield the closed form limiting distribution,

$$\mathcal{L}_0(2L) \to \frac{1}{2}\mathcal{L}(\chi_1^2) + \lambda(\phi^*)\mathcal{L}(\chi_2^2) + \left(\frac{1}{2} - \lambda(\phi^*)\right)\mathcal{L}(0).$$

where

$$\lambda(\phi^*) = \left[\arctan\left(1/\sqrt{2}\right) - \arctan\left(-\phi^*/\sqrt{2}\right)\right]/2\pi$$

is the angle between $\phi = \phi^*$ and $\phi = -1$ after a normalizing transformation of the (θ_1, θ_2) space.

5.6 Limiting behavior as $k \to \infty$.

In the limit as $n \to \infty$, we deal with the supremum of a Gaussian process with mean 0, variance 1 and the autocorrelation function $\rho(\phi_1, \phi_2)$ in (4.11). We shall show that as k gets large neighboring values of the stochastic process becomes almost independent and hence the supremum resembles that of many $i.i.d.$ $N(0,1)$ random variables and approaches infinity. As we shall show, this approach is of the order of $(\log k)^{1/2}$ which grows very slowly with k.

To demonstrate the almost independence, consider ϕ_1 and $\phi_2 = \phi_1 + \delta$ where ϕ_1, is bounded away from 0, $k\delta^2 \to \infty$ while $k\delta^3 \to 0$, e.g. $k = \delta^{-2/3}$. Then a careful expansion yields

$$\ell n\, \rho^2(\phi_1, \phi_2) = \frac{-k\delta^2}{(1+\phi_1^2)^2} + o(1) \to -\infty$$

and $\rho(\phi_1, \phi_1 + \delta) \to 0$.

This analysis suggests the transformation $u = -k^{1/2}\phi$ for $-1 \le \phi \le 0$, so that as $k \to \infty$

$$\rho(\phi_1, \phi_2) = \frac{(1+\phi_1\phi_2)^k - 1}{\{[(1+\phi_1^2)^k - 1][(1+\phi_2^2)^k - 1]\}^{1/2}} \to \frac{e^{u_1 u_2} - 1}{[(e^{u_1^2} - 1)(e^{u_2^2} - 1)]^{1/2}} = \rho_1(u_1, u_2)$$

where

$$\rho_1(u_1 + a, u_2 + a) \to \rho_2(u_1 - u_2) = e^{-\frac{1}{2}(u_1 - u_2)^2} \tag{5.2}$$

as $a \to \infty$. Our stochastic process converges in law to one with autocorrelation function ρ_1. Here ρ_2 is the autocorrelation function of a stationary Gaussian stochastic process for which the asymptotic properties of the supremum are well known. In Appendix 4 we show that this supremum, which, over the interval $0 \le u \le k^{1/2}$, is $(\log k)^{1/2} + O_p(1)(\log k)^{-1/2}$, dominates stochastically that of our limiting process determined by ρ_1.

Two points are worth noting. First, with large probability the supremum for the process corresponding to ρ_1 takes place for large u, and is, in the limit, stochastically equal to that for ρ_2. Second, our asymptotics involves a double limit. First we let $n \to \infty$

and then $k \to \infty$. In practice both k and n are finite, and for large k, very large n is required for the asymptotic normality to be meaningful for ϕ somewhat distant from 0, i.e., p removed from 0.5. We shall not elaborate on this point except to remark that informal calculations suggest that our approximations through the asymptotic theory lead to estimates of the quantiles of $2L$ that are conservative in the following sense. They are somewhat larger than for finite k and n and their use would suggest larger P values than the true values.

6. Simulations

In this section we compare various asymptotic and finite sample distributions for twice the logarithm of the likelihood ratio. With a few exceptions for $k = 2$, the asymptotic distributions are not expressed in simple closed form and hence were calculated by simulation. We present various estimated quantiles for each distribution. The calculations were based on 10,000 simulations each, and so the standard deviation of the estimate $\hat{x}_q$ of the quantile x_q would be $(0.01)[q(1-q)/f^2(x_q)]^{1/2}$. Since the density of $f(x_q)$ is ordinarily not known, an interested user of the tables could crudely estimate it from the table. Since the finite sample distributions are actually discrete, there is some indication of *granularity* for small n and q close to one. Table 1 does not reveal this granularity and its effect on the standard deviation of our estimated quantiles, for which coarse approximations can be inferred from the limited data presented here. From a sample of 10,000,only crude results can be expected for estimating the 0.999 quantile based on the 10 largest observations, an estimate which usually has a relatively large standard deviation.

Table 1 presents the quantiles and estimated quantiles in the following order. First the asymptotic case $(n = \infty)$ for the one sided problem $(s = 1,\ 0 < p \leq .5)$ is tabulated. Then a couple of cases for the two sided problem $(s = 2)$ are listed. Thereafter all entries correspond to the one sided case. The finite sample results are listed. A few examples of the mixed case with two values of k are given. In the asymptotic version the value of n

is replaced by the ratio λ_1/λ_2. In the finite sample version n_1, n_2 is used for n and in both cases $k_1 : k_2$ is given for k.

The noncentral asymptotic case lists the value of p and $\theta_1^* = \sqrt{n}\theta_1$. For the finite sample versions, the corresponding values of α are also listed. Then the restricted case is tabulated in the asymptotic and finite sample size case. The table concludes with the quantiles for chi-square with one and two degrees of freedom and $(.01)[q(1-q)]^{1/2}$ to help compute standard deviations.

The entries in Table 1 were culled from a more extensive list of simulations to provide the reader with some ability to make meaningful comparisons without the benefit of sensory overload.

The asymptotic results of Table 1 show that $2L$ grows slowly with k. For $k = 40, 2L$ for the one sided problem is still stochastically smaller than chi-square with 2 degrees of freedom. For the two sided problem $2L$ tends to be little larger, which is to be expected. The finite sample estimates give lower quantiles than the asymptotic values. Thus the asymptotic values are conservative in the sense that they would lead to overestimating the P values of the test. The values of k and n in our simulations are not sufficiently large to exhibit the asymptotic behavior for large k except in the crudest fashion.

Except for the unreliable 0.999 estimated quantile, the mixed distribution is remarkably insensitive to the mixture proportions, yielding quantiles close to those of the unmixed larger k values.

For the noncentral asymptotic distributions a difference in the value of p seems to lead to an effect which is largely that of a translation. For $k = 2$, the asymptotic distribution is a good approximation for sample sizes of $n = 20$ and 40. However for $k = 10$, the asymptotic distribution has $2L$ much larger stochastically than it should be for the sample sizes $n = 20$ and 40. This effect, which suggests an unduly optimistic estimate of the power of the test, is due to the fact that the central limit theorem takes effect rather slowly when k is substantial and p is far from .5. Then $-\phi$ is relatively large and $S(Z,\phi)$ tends to have a large kurtosis. Presumably this approximation would be good for very large sample sizes, much larger than appropriate for our genetic application when $k=10$.

Finally, for the restricted problem, and the values of p, the asymptotic distribution (under the hypothesis) seems to be as expected, stochastically smaller than that for the unrestricted case. However this reduction, in not large and becomes a little larger as the sample size diminished.

Table 2 presents $K(\theta_0,0)$ and $V(\theta_0)^{1/2}$ which can be used with a central limit theorem approximation to estimate the quantiles for the finite sample version of the noncentral distribution. In fact the highly right skewed distribution of $2L$ is such that the normal approximation is not very good in the right tails. In the more crucial left tails, it is much better, but still leads to *conservative* estimates of the power function by overestimating the error probability of tests for the noncentral case. This overestimate is due to the natural truncation of the real distribution of $2L$ at 0.

APPENDIX 1. Details on the case $k=2$.

If we introduce the transformation $\theta_1^* = \sqrt{32}\theta_1$ and $\theta_2^* = 4\theta_2$, then the unrestricted MLE $\hat{\theta}_u^*$ of θ^* satisfies

$$\mathcal{L}_0(\sqrt{n}\hat{\theta}_u^*) \to N(0,1)$$

but the set of possible values of θ^* is the union of the region, between $\theta_2^* = \theta_1^*/\sqrt{2}$ and $\theta_2^* = \theta_1^{*2}/2$ for $0 \le \theta_1^* \le \sqrt{2}$, and its reflection about the θ_2^* axis. In the neighborhood of $\theta^* = 0$, this set is approximated by the region, between $\theta_2^* = \theta_1^*/\sqrt{2}$ and $\theta_2^* = 0$ for $\theta_1^* > 0$, and it reflection. Then $2L$ behaves like the difference between the squared distance of $\sqrt{n}\hat{\theta}_u^*$ to the origin and to the above set. When $\sqrt{n}\hat{\theta}_{2u}^* < 0$, this difference is $(\sqrt{n}\theta_1^*)^2$ and contributes $\frac{1}{2}\mathcal{L}(\chi_1^2)$ to the asymptotic distribution of $\mathcal{L}_0(2L)$. Here $\hat{\theta}_2^* \approx 0$ and $\hat{\theta}_1^* \approx \hat{\theta}_{1u}^*$. When $\hat{\theta}_u^*$ is in the restricted region $\hat{\theta}^* = \hat{\theta}_u^*$

and the squared distance to the origin is $\hat{\theta}_{1u}^{*2} + \hat{\theta}_{2u}^{*2}$, this contributes χ_2^2 with probability determined by the proportion of the region covered by the restricted region, i.e., 2λ where $\lambda = (2\pi)^{-1}\arctan(1/\sqrt{2})$. In the remaining region, the squared distance is the squared length of the projection of $\sqrt{n}\hat{\theta}_u^*$ onto the nearer of the two lines $\theta_2^* = \pm\theta_1^*/\sqrt{2}$. Since the projections of $\sqrt{n}\hat{\theta}_u^*$ onto $\theta_2^* = \theta_1^*/\sqrt{2}$ and onto the direction vertical to it are independent $N(0,1)$ random variables, we can represent this contribution by

$$\left(\frac{1}{2} - 2\lambda\right)\mathcal{L}\left(Y_1^2 | 0 < Y_2 < \sqrt{2}Y_1\right).$$

Equations (3.3) - (3.5) follow readily. For Problem 3, a similar but somewhat simpler analysis yields Equation (3.6) after it is noted that for a proportion $(1/2 - \lambda)$ of the region, the closest point from $\sqrt{n}\hat{\theta}_u^*$ to the restricted region is at the origin.

APPENDIX 2. Estimation of the parameters by the method of moments.

We estimate the parameters of Problem 1 by matching the first two sample moments with the theoretical moments. We calculate

$$\begin{aligned}
\mu_1 &= E(X) = \alpha kp + (1-\alpha)k/2 = k(2\theta_1 + 1/2) \\
\mu_2 &= E(X^2) = k\left[\alpha p + (1-\alpha)/2\right] + k(k-1)\left[\alpha p^2 + (1-\alpha)/4\right] \\
&= k(2\theta_1 + 1/2) + k(k-1)(\theta_1\phi + 2\theta_1 + 1/4) \\
\theta_1 &= \frac{1}{2}\left[\frac{\mu_1}{k} - \frac{1}{2}\right] \\
\theta_2 &= \left[\frac{\mu_2 - \mu_1}{k(k-1)} - \frac{\mu_1}{k} + \frac{1}{4}\right]
\end{aligned}$$

By the Central Limit Theorem

$$\begin{aligned}
\overline{X} = \frac{1}{n}\Sigma X_i &= \mu_1 + \frac{Z_1}{\sqrt{n}} \\
\frac{1}{n}\Sigma X_i^2 &= \mu_2 + \frac{Z_2}{\sqrt{n}}
\end{aligned}$$

where $\mathcal{L}(Z_1, Z_2) \to N(0, \Sigma^*), \Sigma^* = \|\sigma^*_{ij}\|$, and $\sigma^*_{ij} = \mu_{i+j} - \mu_i\mu_j$ with $\mu_i = E(X^i)$. Under the hypothesis H_o, $\mathcal{L}(X) = Bi(k, 1/2)$, and $\mu_1 = k/2$, $\mu_2 = (k + k^2)/4$, $\mu_3 = k^2(k+3)/8$ and $\mu_4 = k(k^3 + 6k^2 + 3k - 2)/16$. Then $\sigma^*_{11} = k/4$, $\sigma^*_{12} = k^2/4$, and $\sigma^*_{22} = k(k+1)(2k-1)/8$. Applying the method of moments and ignoring the restrictions on θ_1 and θ_2 yields the estimates

$$\tilde{\theta}_{1u} = \frac{1}{2}\left[\frac{\overline{X}}{k} - \frac{1}{2}\right] = \frac{1}{2k} Z_1 n^{-1/2}$$

and

$$\tilde{\theta}_{2u} = \frac{\frac{1}{n}\Sigma X_i^2 - \overline{X}}{k(k-1)} - \frac{\overline{X}}{k} + \frac{1}{4} = \frac{1}{k(k-1)}(Z_2 - Kz_1)n^{-1/2}$$

where $E\{Z_1(Z_2 - Kz_1)\} = \sigma^*_{12} - k\sigma^*_{11} = 0$ and hence

$$\mathcal{L}(\sqrt{n}\tilde{\theta}_u) \to N(0, \Sigma_u)$$

with Σ_u a diagonal matrix with diagonal elements $\sigma^*_{11}/4k^2 = (16k)^{-1}$ and $(\sigma^*_{22} - k\sigma^*_{12})/k^2(k-1)^2 = [8k(k-1)]^{-1}$. Thus $\tilde{\phi}_u = \tilde{\theta}_{2u}/\tilde{\theta}_{1u}$ has a limiting Cauchy Distribution with scale parameter σ where $\sigma^2 = 2(k-1)^{-1}$.

APPENDIX 3. The noncentral case.

Let the true value of θ be $\theta_{0n} = n^{-1/2}(\theta^*_1, \theta^*_2)$ where $\theta^*_2 = \phi^*\theta^*_1$. We recapitulate the derivation of Section 4 for this case. We have

$$M_j = nf(j, \theta_{0n}) + \sqrt{n}Z_j$$

where once again, as in (4.4), $\mathcal{L}_{\theta_{0n}}(Z) \to N(0, \Sigma)$ as $n \to \infty$. Then, for $\theta = 0(n^{-1/2})$,

$$\begin{aligned} L(\theta) &= \sum_j (nf(j, \theta_{0n}) + \sqrt{n}Z_j)\ell n(1 + u(j, \theta)) \\ &= \sum \left\{ nf(j, \theta_{0n})\left[u(j, \theta) - \frac{1}{2}u^2(j, \theta)\right] + \sqrt{n}Z_j u(j, \theta)\right\} + o_p(1). \end{aligned}$$

The first part of the above expression being summed is

$$A_j = n2^{-k}\binom{k}{j}\left\{1+\frac{4\theta_1^*}{\sqrt{n}}P_{kj}(\phi^*)\right\}\left\{4\theta_1 P_{kj}(\phi) - 8\theta_1^2 P_{kj}^2(\phi)\right\},$$

the sum of which can be expressed in terms of sums of the form

$$\begin{aligned} a_{rs} &= \sum 2^{-2}\binom{k}{j}(1+\phi^*)^{rj}(1+\phi)^{sj}(1-\phi^*)^{k-j}(1-\phi)^{s(k-j)} \\ &= \left\{[(1+\phi^*)^r(1+\phi)^s + (1-\phi^*)^r(1-\phi)^s]/2\right\}^k \end{aligned}$$

where $r = 0$ or 1 and $s = 0, 1,$ or 2. The two terms of A_j without θ_1^* were already summed in the previous derivation and contribute $-8n\theta_1^2 P_{kk}(\phi^2)$. The remaining terms yield

$$16\theta_1^*\theta_1\sqrt{n}\left[\frac{a_{11}-a_{10}-a_{01}+a_{00}}{\phi^*\phi}\right] - 32\theta_1^*\theta_1^2\sqrt{n}\left[\frac{(a_{12}-a_{02})-2(a_{11}-a_{01})+(a_{10}-a_{00})}{\phi^*\phi^2}\right]$$

But $a_{00} = a_{10} = a_{01} = 1$ and with some calculation we have

$$\sqrt{n}\{16\theta_1^*\theta_1 P_{kk}(\phi^*\phi) - 32\theta_1^*\theta_1^2 Q_k(\phi^*, \phi)\}$$

where

$$Q_k(\phi^*, \phi) = \{[(1+\phi^2+2\phi\phi^*)^k - (1+\phi^2)^k] - 2[(1+\phi^*\phi)^k - 1]\}/\phi^*\phi^2.$$

For fixed ϕ we maximize, with respect to θ, subject to $\theta_1\phi > 0$, the main term of

$$L(\phi) = -\theta_1^2\left\{8nP_{kk}(\phi^2) + 32\theta_1^*\sqrt{n}Q_k(\phi^*, \phi)\right\} + 4\theta_1\sqrt{n}S^*(Z, \phi) + o_p(1)$$

where

$$\begin{aligned} S^*(Z,\phi) &= S(Z,\phi) + 4\theta_1^* P_{kk}(\phi^*\phi) \\ &= \sum W_j\phi^{j-1} + 4\theta_1^*\binom{k}{j}(\phi^*\phi)^{j-1} = \sum W_j^*\phi^{j-1} \end{aligned}$$

and the relation of W to Z is the same as in (4.10) and the limiting distribution of W^* is that of independent normals with

$$\lim \mathcal{L}_{\theta_{0n}}(W_j^*) = N\left(4\theta_1^* \binom{k}{j} \phi^{*j-1}, \binom{k}{j}\right)$$

The maximum of the main term of $L(\phi)$ occurs at $\theta_1 = 0$ or at

$$\theta_1 = \frac{S^*(Z,\phi)}{4\sqrt{n}P_{kk}(\phi^2) + O(1)}$$

if $S^*(Z,\phi)$ has the same sign as ϕ. Then

$$2L(\phi) = T^{*2}(\phi) + o_p(1)$$

and Theorem 2 follows.

APPENDIX 4. The case of $k \to \infty$.

We shall show that ρ_1 and ρ_2 defined in Section 5.6 satisfy $\rho_1(u_1,u_2) \geq \rho_2(u_1 - u_2)$. Hence Slepian's Lemma in Leadbetter et al. (1982, p. 156) implies that the supremum of the Gaussian process with covariance function ρ_1 is stochastically less than that of the stationary Gaussian process with autocovariance function ρ_2. Then, applying Theorem 8.2.7 in Leadbetter et al. (1982, p. 171) for the stationary process with $\lambda_2 = -\rho_2''(0) = 1$, we have the supremum over the range $0 \leq u \leq \sqrt{k}$ to be

$$M = (\log k)^{1/2} + \frac{X - \log(2\pi)}{(\log k)^{1/2}} + o_p(\log k)^{-1/2}$$

where

$$P\{X < x\} = \exp(-e^{-x}).$$

We conclude with a proof of

LEMMA 2. $\rho_1(u_1,u_2) \geq \rho_2(u_1 - u_2)$ for $0 \leq u_1,\ 0 \leq u_2$.

Proof: First we note that $\rho_1 \geq 0$ and that

$$\rho_1^2(u_1,u_2) = \rho_2^2(u_1 - u_2)[1 - e^{-u_1u_2}]^2/[1 - e^{-u_1^2}][1 - e^{-u_2^2}].$$

Hence it suffices to prove that

$$(1-e^{-u_1u_2})^2 \geq (1-e^{-u_1^2})(1-e^{-u_2^2})$$

or

$$\log(1-e^{-u_1u_2}) \geq \frac{1}{2}\log(1-e^{-u_1^2}) + \frac{1}{2}\log(1-e^{u_2^2})$$

If we let $u = \exp(y)$, it suffices to show that

$$g(y) = \log h(y) = \log[1-\exp(-\exp(2y))]$$

is a concave function of y or that $g''(y) \leq 0$. We calculate

$$h'(y) = 2\exp(2y)\exp[-\exp(2y)] > 0$$

$$h''(y) = 2h'(y)[1-\exp(2y)]$$

$$g'(y) = h'(y)/h(y)$$

and

$$g''(y) = -\left[\frac{h'(y)}{h(y)}\right]^2 + \frac{h''(y)}{h(y)}$$

If $y > 0$, $h''(y) < 0$ and $g''(y)$ is clearly negative. For $y < 0$, we must show that

$$[h'(y)]^2 \geq h(y)h''(y)$$

or

$$2\exp(2y)\exp(-\exp(2y)) \geq 2(1-\exp(2y))(1-\exp(-\exp(2y))$$

$$\exp(2y) + \exp(-\exp(2y)) \geq 1$$

Let $v = \exp(2y)$. Then we need

$$e^{-v} + v - 1 \geq 0 \qquad \text{for} \qquad 0 \leq v \leq 1.$$

In fact it is well know that the above inequality holds for all v.

Bibliography

[1] Chernoff, H. (1954). "On the distribution of the likelihood ratio", *Ann. Math. Statist.*, **25**, 573-578.

[2] Ghosh, J. K. and Sen, P.K. (1985). "On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results," Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, edited by L. M. LeCam and R. A. Olshen, Wadsworth and Inst. of Math. Statist., Voll. II, pp. 789-806.

[3] Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1982), *Extremes and Related Properties of Random Sequences and Processes*, New York, Springer-Verlag.

[4] Wilks, S. S. (1938). "The large sample distribution of the likelihood ratio for testing composite hypothesis," *Ann. Math. Statist.*, **9**, 60-62.

Table 1. Quantiles of $\ell n(2L)$

q	0.25	0.50	0.75	0.90	0.95	0.99	.999	k	n	s	p	$-\theta_1^*$	α	p^*
asymptotic;	0	.06	.80	2.2	3.4	6.3	10.7	2	∞	1				
	0	.17	1.07	2.5	3.7	6.5	10.8	4	∞	1				
	0	.26	1.26	2.9	4.3	7.2	11.2	6	∞	1				
	.01	.42	1.56	3.2	4.5	7.4	12.0	10	∞	1				
	.30	1.12	2.54	4.3	5.8	8.9	13.4	40	∞	1				
asymptotic; two-sides	.23	.77	1.85	3.4	4.6	7.6	12.0	2	∞	2				
	.56	1.34	2.75	4.5	5.8	9.1	14.8	10	∞	2				
finite sample;	0	0	.6	2.0	3.1	5.9	8.6	2	40	1				
	0	0	.8	2.0	3.5	5.8	8.6	2	20	1				
	0	.1	.9	3.0	4.4	7.0	11.8	10	40	1				
	0	0	.7	2.4	3.8	6.5	10.1	10	20	1				
	0	.1	.9	2.4	3.6	6.8	11.7	40	40	1				
	0	0	.8	2.2	3.4	6.4	11.6	40	20	1				
mixed; asymptotic;	0	.4	1.6	3.3	4.5	7.8	14.4	3:10	1/3	1				
	0	.4	1.6	3.2	4.5	7.9	11.9	3:10	3/1	1				
mixed; finite sample;	0	.1	.8	2.4	4.2	6.5	13.7	3:10	10,30					
	0	.1	.8	2.3	3.5	6.7	11.6	3:10	30,10					
noncentral; asymptotic;	.5	1.8	4.0	6.8	8.7	13.3	19.4	2	∞	1	.30	.2		
	.6	2.1	4.4	7.4	9.4	13.9	22.7	2	∞	1	.15	.2		
	3.3	6.2	10.0	14.4	17.2	23.7	31.9	2	∞	1	.30	.4		
	3.9	7.1	11.1	15.4	18.2	25.3	33.4	2	∞	1	.15	.4		
	10.1	14.6	20.2	25.9	29.5	36.7	47.6	10	∞	1	.30	.2		
	59.5	70.3	82.0	92.7	100	114	128	10	∞	1	.15	.2		
	45.1	54.8	65.4	75.9	82.5	96	111	10	∞	1	.30	.4		
	254	277	300	322	336	363	396	10	∞	1	.15	.4		
noncentral; finite sample;	.5	1.7	3.9	6.4	8.8	12.9	19.3	2	40	1	.30	.2	.32	
	.3	1.4	3.7	6.3	8.3	12.1	18.2	2	20	1	.30	.2	.45	
	.5	1.9	4.2	7.4	9.3	13.8	21.6	2	40	1	.15	.2	.18	
	.3	1.7	3.8	6.6	8.4	13.5	18.4	2	20	1	.15	.2	.26	
	3.1	5.8	8.4	13.4	16.4	21.9	29.2	2	40	1	.30	.4	.63	
	2.6	5.8	8.6	12.6	15.4	20.5	28.3	2	20	1	.30	.4	89	
	3.1	6.0	9.9	13.9	17.1	23.8	30.6	2	40	1	.15	.4	.36	
	3.5	5.9	9.9	14.2	16.4	22.2	31.0	2	20	1	.15	.4	.51	

Table 1. Quantiles of $\ell n(2L)$ cont.

q	0.25	0.50	0.75	0.90	0.95	0.99	.999	k	n	s	p	$-\theta_1^*$	α	p^*
	5.1	9.1	14.4	19.9	23.6	30.8	43.7	10	40	1	.30	.2	.32	
	4.8	8.4	13.4	18.7	22.7	30.5	40.4	10	20	1	.30	.2	.45	
	9.9	16.5	24.5	33.4	39.2	52.1	68.9	10	40	1	.15	.2	.18	
	7.7	14.4	21.7	30.3	35.7	47.7	60.6	10	20	1	.15	.2	.26	
	21.5	29.0	37.7	46.0	51.6	62.7	78.1	10	40	1	.30	.4	.63	
	20.5	27.1	34.6	42.1	46.8	57.8	72.0	10	20	1	.30	.4	.89	
	33.7	45.3	58.6	72.0	80.8	97.8	114	10	40	1	.15	.4	.36	
	27.6	38.0	50.2	61.6	69.0	84.1	98.7	10	20	1	.15	.4	.51	
	89	111	137	161	176	206	245	20	40	1	.15	.4	.36	
	156	192	232	267	291	331	383	40	20	1	.15	.4	.51	
restricted;														
asymptotic;	0	.35	.72	2.1	3.3	6.2	10.5	2	∞	1				.60
	0	.12	.61	1.9	3.1	5.9	10.1	2	∞	1				.75
	0	.32	1.4	3.0	4.3	7.6	12.9	10	∞	1				.60
	0	.12	.97	2.5	3.7	6.9	10.8	10	∞	1				.75
	.17	.86	2.2	4.0	5.3	8.5	12.3	40	∞	1				.60
	0	.40	1.5	3.2	4.5	8.0	12.3	40	∞	1				.75
restricted;														
finite sample;	0	0	.6	2.0	3.1	6.6	10.8	2	40	1				.60
	0	0	.8	2.0	3.6	6.0	10.1	2	20	1				.75
	0	0	.5	2.0	3.0	5.7	13.0	2	40	1				.60
	0	0	.5	2.0	2.9	5.9	9.4	2	20	1				.75
	0	0	.1	1.5	2.9	6.6	12.0	40	40	1				.60
	0	0	.1	1.4	2.5	6.0	10.3	40	20	1				.75
	0	0	0	0	.3	3.2	7.4	40	40	1				.60
	0	0	0	0	.6	4.0	8.5	40	20	1				.75
chi-square;														
with 1 df	.10	.46	1.3	2.7	3.8	6.6	10.8							
with 2 df	.58	1.39	2.8	4.6	6.0	9.2	13.8							
$(.01)[q(1-q)]^{1/2}$	.004	.005	.004	.003	.002	.001	.0004							

Table 2

Kullback-Leibler Information $K(\theta_0, 0)$ and
Standard Deviation $SD = [V(\theta_0)]^{1/2}$ of $2L$

k									
	p	.300	.300	.300	.300	.150	.150	.150	.150
	$-\theta_1$	.032	.045	.063	.089	.032	.045	.063	.089
2	KL	.017	.033	.066	.131	.019	.037	.072	.141
2	SD	.185	.260	.362	.499	.198	.278	.389	.537
6	KL	.057	.108	.205	.395	.092	.163	.288	.51
6	SD	.364	.502	.682	.892	.511	.691	.921	1.20
10	KL	.105	.193	.354	.66	.204	.34	.57	.94
10	SD	.526	.711	.942	1.18	.880	1.14	1.46	1.82
20	KL	.253	.44	.76	1.34	.58	.90	1.40	2.17
20	SD	.924	1.20	1.51	1.77	1.88	2.32	2.81	3.30
40	KL	.63	1.01	1.64	2.73	1.50	2.21	3.27	4.85
40	SD	1.72	2.12	2.52	2.69	3.94	4.68	5.45	6.09

For comparisons with the noncentral entries of Table 1
we note θ_1^* and α for the following values of n

n	40	20	40	20	40	20	40	20
$-\theta_1^*$	.200	.200	.400	.400	.200	.200	.400	.400
α	.316	.447	.632	.894	.181	.256	.361	.511

ABSTRACT

A problem of interest in genetics is that of testing whether a mixture of two binomial distributions $B_i(k,p)$ and $B_i(k,1/2)$ is simply the pure distribution $B_i(k,1/2)$. This problem arises in determining whether we have a genetic marker for a gene responsible for a *heterogeneous trait*, that is a trait which is caused by any one of several genes. In that event we would have a nontrivial mixture involving $0 < p < 0.5$ where p is a recombination probability.

Standard asymptotic theory breaks down for such problems which belong to a class of problems where a *natural* parameterization represents a single distribution, under the hypothesis to be tested, by infinitely many possible parameter points. That difficulty may be eliminated by a transformation of parameters. But in that case a second problem appears. The regularity conditions demanded by the applicability of the Fisher Information fails when $k > 2$. We present an approach where use is made of the Kullback Leibler information, of which the Fisher information is a limiting case.

Several versions of the binomial mixture problem are studied. The asymptotic analysis is supplemented by the results of simulations. It is shown that as $n \to \infty$, the asymptotic distribution of twice the logarithm of the likelihood ratio corresponds to the square of the supremum of a Gaussian stochastic process with mean 0, variance 1 and a well behaved covariance function. As $k \to \infty$ this limiting distribution grows stochastically as the square root of $\log k$.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR No. ONR-C-7	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Dept. of Statistics Harvard University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics SC713 Harvard University Cambridge, MA 02138	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION ONR 8b. OFFICE SYMBOL (If applicable) Code 1111	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-91-J1005
8c. ADDRESS (City, State, and ZIP Code) Office of Naval Research Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Asymptotic Distribution of the Likelihood Ratio Test that a Mixture of Two Binomials is a Single Binomial

12. PERSONAL AUTHOR(S)
Herman Chernoff and Eric Lander

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM ____ TO ____	May 17, 1991	24

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
asymptotic distribution, likelihood ratio, extreme values, mixture, genetic markers, heterogeneous traits

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See reverse side.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Herman Chernoff	22b. TELEPHONE (Include Area Code) 617-495-5462 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED